



**Universitat  
Pompeu Fabra**  
*Barcelona*

Department  
of Economics and Business

**Economics Working Paper Series**

**Working Paper No. 1593**

**Experiment-inspired comments on John  
Roemer's theory of cooperation**

**Antoni Bosch-Domènech and Joaquim Silvestre**

**December 2017**

December 15, 2017

**EXPERIMENT-INSPIRED COMMENTS ON JOHN ROEMER'S THEORY OF COOPERATION\***

by

Antoni Bosch-Domènech,  
*Universitat Pompeu Fabra*

and

Joaquim Silvestre  
University of California, Davis

Prepared for the December 18, 2017, Roundtable in honor of John E. Roemer  
Queen Mary University of London

**INDEX**

1. Introduction
2. A cooperative species
  - 2.1. Dominant strategy: The temperature experiment
  - 2.2. A social situation
  - 2.3. Experimental results
  - 2.4. Motivations of the experimental cooperators
3. Towards a theory of cooperation
  - 3.1. Microfoundations
  - 3.2. Rationality
    - 3.2.1. The decision-relevant utility function
    - 3.2.2. Cournot/Nash behavior
    - 3.2.3. Roemer simple Kantian behavior
  - 3.3. Equilibrium
4. Roemer Simple Kantian Equilibrium as a theory of cooperation
  - 4.1. Solidarity
  - 4.2. Trust
  - 4.3. Symmetry
  - 4.4. Monotonicity
5. Conclusion

---

\* We thank John Roemer and Rosemarie Nagel for comments, with the usual *caveat*. We acknowledge the assistance of Pablo López-Aguilar Beltrán, Manager of the BES\_Lab, and financial support from the Spanish *Ministerio de Economía, Industria y Competitividad*, grant MINECO-ECO2014-56154 P.

## 1. Introduction

Motivated by Roemer’s Kantian equilibrium (John Roemer, 2010, 2015, In Press) we performed some experiments on cooperation in extremely simple treatments (Bosch-Domènech and Silvestre, 2017, and Appendix below). Here we comment on some lessons that we have learned from them, and offer some thoughts on Roemer’s theory of cooperation based on social situations that can be easily compared to those in our experiments.

## 2. A cooperative species

### 2.1. Dominant strategy: The temperature experiment

Roemer (In Press) emphasizes the extent of cooperation among humans, particularly relative to our great ape cousins. We contribute to the extensive experimental literature on cooperation with treatments where the dominant strategy in the experimental payoffs prescribes not to cooperate, and compare two “strategically isomorphic” situations of decision making: in one, the payoffs depend on a parameter given by nature; in the other, on the decisions of other participants.

The Appendix below describes a one-shot experimental treatment of the first kind with the following payoff matrix.

|                   | Temperature in Istanbul $\leq 16^{\circ}\text{C}$ | Temperature in Istanbul $> 16^{\circ}\text{C}$ |
|-------------------|---|--|
| You choose CIRCLE | You earn €10                                      | You earn €1                                    |
| You choose SQUARE | You earn €15                                      | You earn €6                                    |

Table 1. Payoff matrix in the Temperature Treatment

Clearly, SQUARE is the Dominant Strategy in the experimental payoffs. Indeed, all twenty participants in our experiment chose SQUARE, and seventeen of them justified their decision by referring to the dominant-strategy property, i. e., by choosing SQUARE my earnings are higher than if I chose CIRCLE, *no matter what the temperature is in Istanbul*.

### 2.2. A social situation

One of the experimental treatments in Bosch-Domènech and Silvestre (2017), which we label the *2-Person, Frameless PD Treatment*, is based on a Prisoner’s Dilemma game. There, each participant faces the payoff matrix of the Temperature Treatment of Table 1, but, instead of the temperature in Istanbul, the columns correspond to the choice of

another player (randomly and anonymously selected). Accordingly, the payoff matrix is given by Table 2. The participants were informed that the payoffs of the other player were symmetrical to theirs and, in fact, they were explicitly given the payoff matrix of the other player.

|                   |                                 |                                 |
|-------------------|---------------------------------|---------------------------------|
|                   | The other player chooses CIRCLE | The other player chooses SQUARE |
| You choose CIRCLE | You earn €10                    | You earn €1                     |
| You choose SQUARE | You earn €15                    | You earn €6                     |

Table 2. Payoff matrix in the 2-Person, Frameless PD Treatment of Bosch-Domènech and Silvestre (2017)

Table 3 gives, for future reference, the payoff bimatrix of the game defined by Table 2.

|            |        |               |           |
|------------|--------|---------------|-----------|
|            |        | Column Player |           |
|            |        | CIRCLE        | SQUARE    |
| Row Player | CIRCLE | €10<br>€10    | €1<br>€15 |
|            | SQUARE | €1<br>€15     | €6<br>€6  |

Table 3. Payoff bimatrix of a Prisoner's Dilemma Game

Tables 1 and 2 have identical payoffs to a participant, and the choices open to her are similarly labeled CIRCLE and SQUARE. In this sense, the two situations are “strategically isomorphic.” The main difference is the labeling of columns: impersonal states of nature vs. the choices of a human, anonymous decision maker. In other words, the experimental treatment depicted in Table 2 is *social*, whereas that of Table 1 is not.

### 2.3. Experimental results

Twenty eight percent of the participants (13/46) in the 2-Person Frameless PG Treatment choose CIRCLE, in stark contrast with the 0% (0/20) who did so in the Temperature Treatment. Thus, our experimental results of the simple treatments of Tables 1 and 2 yield a

powerful conclusion: many decision makers act differently when, *ceteris paribus*, the outcomes of their decisions depend *on an impersonal, natural variable vs. the decisions of other human beings*.

As we argue in the following section, it is not unreasonably to assume that the participants in the treatment of Table 2 who choose CIRCLE are acting rationally. But how can such rationality be justified? Indeed, in the “isomorphic” game of Table 1 any sensible notion of rationality induces the choice of SQUARE.

#### 2.4. Motivations of the experimental cooperators

In addition to the Treatment described in Section 2.2 above, Bosch-Domènech and Silvestre (2017) also reports on several other treatments. There we find that, in social situations, nonnegligible numbers of participants choose a cooperative action which is strategically dominated in terms of the experimental payoffs. What are their motivations?

Ralph Bayer *et al.* (2013), Maxwell Burton-Chelley *et al.* (2016), Burton-Chelley and West (2013), and Daniel Houser and Robert Kurzban (2002) have found a correlation between *confusion* and cooperation in lab experiments. But it is hard to believe that the game of Table 2 has a greater potential for confusion than that of Table 1.

On the other hand, the literature has documented *experimenter effects* as described by David Cooper and John Kagel (2015): experimental subjects are “... trying to figure out what the experimenter wants and then doing it to please him or her.” We argue in Bosch-Domènech and Silvestre (2017) that the design of our treatments makes it unlikely that decisions of our cooperators are notably influenced by experimenter effects.

All our treatments share two simplifying features. First, they are one-shot. This discourages the *strategic cooperation* typical of the early stages of repeated games and reduces the possible motivation of *reciprocity*, emphasized by Samuel Bowles and Herbert Gintis (2011) as a foundation for cooperation (see also Kenneth Clark and Martin Sefton, 2001). Second, our participants had to choose between (at most) two options. This neutralizes a potential tendency of choosing something away from the extremes, which has been argued to explain a degree of cooperation, *e. g.*, by Rolf Kümmerli *et al.* (2010).

We asked participants to explain the reasons for their decisions, both in open-format and in closed-format questionnaires, and we paid special attention of how the participants who chose to cooperate (or CIRCLE) explained their decisions. The closed-list format consisted of the following options.

1. I chose at random;

2. It is the choice that I'd like everybody to make in this situation;
3. I like to help others even at a cost to myself;
4. In this manner, our joint earnings are higher;
5. Taking advantage of others is not right.

The most frequent response was the fourth one, indicating a concern for *efficiency*. Next came number two, of *Kantian* flavor. *Altruism* (number three) and *fairness* were adopted by just a few. (Nobody chose 1.)

As just noted, in the nonsocial situations our participants universally choose SQUARE, the payoff-dominant strategy. Moreover, they explicitly justify their choice in terms of the dominant-strategy property.

But in the social treatments they do not refer to the dominant-strategy feature of the payoff dominant strategy. We conjecture that they do not consider this property important enough because they focus on something else.

They realize that everybody would be better off if everybody chose CIRCLE. Let's call this the *right thing to do*. Different people resolve the conflict between payoff dominance and the "right thing to do" in different ways, with frequencies that vary among the features of the various treatments. Participants who choose the "right thing to do" place its value above that of their own individual payoffs.

How is the right thing to do defined? As just noted, we find that our cooperating participants appeal most frequently either by *group efficiency* ("this is the way how we all get more" or by *Kantian reasoning* ("I do what I'd like everybody else to do"), with *altruism* failing to play a relevant role.

### **3. Towards a theory of cooperation**

#### **3.1. Microfoundations**

The Cournot/Nash paradigm of social interaction provides a micro-founded theory of noncooperative social interaction. But, as argued by Roemer (In Press), Bowles and Gintis (2011), and many others, human social interaction often yields cooperative outcomes. A microfounded theory of cooperation should satisfy three desiderata:

- (1) *Rationality*. Individuals should decide rationally.
- (2) *Equilibrium*. The decisions by all agents should be compatible among themselves; in particular, if an agent's decision is based on her conjectures about the decisions of other agents, these conjectures should be correct.

(3) *Efficiency*. A theory of cooperation should single out outcomes that are socially efficient.

The following sections discuss these desiderata in relation to the Cournot/Nash and Kant/Roemer models.

### 3.2. Rationality

#### 3.2.1. The decision-relevant utility function

Individual rationality requires that the choices (decisions) of the decision maker aim at reaching his or her goals.

Consider a Decision Maker named  $i$ ,  $DM_i$  in short. She faces a *parameter vector*  $p_i \in P_i$  ( $P_i$  is the set of parameter vectors that she may face): a parameter is defined by the trait that  $DM_i$  views it as outside her control, i. e., in her perception, she cannot influence the parameters that she faces.

She cannot choose the parameters that she faces, but she must decide among the elements of her *decision set*  $S_i$ , called her “strategy” set in game theory.<sup>1</sup> It might occur that a vector of parameters  $p_i$  restricts the set of decisions available to her to a subset of  $S_i$ , as in the budget set of standard consumer theory. But for the purposes of the present discussion we shall assume for simplicity that no such restrictions are present, and hence that she is free to choose any element of  $S_i$ .

Rationality is defined by two requirements. First, the decision maker must have clear, consistent objectives, defined by a complete and transitive binary relation  $\succsim_i$  on  $S_i \times P_i$ .<sup>2</sup> We disregard the case, irrelevant for our discussion here, where  $\succsim_i$  is not representable by a real-valued function, and assume that  $\succsim_i$  is indeed represented by a function:

$$U_i: S_i \times P_i \rightarrow \mathfrak{R},$$

called *i's decision-relevant utility function*.<sup>3</sup>

Second, the decision maker must optimize, i. e., when she faces the parameter vector  $p_i$ , she chooses  $\bar{s} \in S_i$  satisfying  $U_i(\bar{s}, p_i) \geq U_i(s_i, p_i), \forall s_i \in S_i$ .

This definition of rationality captures the ideas of *consistency* of objectives, as described by  $U_i$  and of *congruence* between decisions and the objectives of the decision maker.<sup>4</sup> But it is general enough to encompass objectives other than narrowly defined self-interested outcomes.

<sup>1</sup> In other contexts,  $S_i$  might be called the set of “choices,” “actions,” “acts” or “alternatives” open to her; in traditional models of decision under risk,  $S_i$  may be referred to as a set of “lotteries.”

<sup>2</sup> See Andreu Mas-Colell *et al.* (1995), Definition 1.B.1.

<sup>3</sup> Of course, if  $U_i$  represents  $\succsim_i$ , so does any increasing transformation of  $U_i$ .

It will be useful to assume that  $DM_i$ 's self-interested preferences on outcomes are represented by a *personal material benefit function*  $PMB_i$ . In an experimental context, we can think of the  $PMB_i$  as the payoffs established by the experimenter.  $DM_i$  may see  $PMB_i$  as a factor in her decision-relevant utility function  $U_i$ , but it may not necessarily be the only consideration: other features of the outcomes, such as their fairness, or efficiency, may also be relevant. Moreover, whereas the  $PMB_i$  is defined on outcomes,  $U_i$  is defined on decisions (cross parameters).<sup>5</sup> For instance, two decisions may produce the same outcome (for a given vector of parameters), yet  $DM_i$  may prefer one because it agrees with a given procedural ethical principle.

### 3.2.2. Cournot/Nash behavior

The analysis of Antoine Augustin Cournot (1838) is based on the absence of cooperation among profit-maximizing firms. Of course, the original Cournot oligopolists could increase their profits if they managed to implicitly or explicitly collude, *but they do not*. Instead, they parametrically treat the sales (or, in the model of Cournot, 1838, chapter 9, the price) of other firms as a natural environment, the way our participants in the experimental treatment of Table 1 above treat the temperature in Istanbul.<sup>6</sup>

Cournot's noncooperative setup is generalized by John Nash (1950, 1951). Again, a Cournot/Nash player views other players' actions as an impersonal natural environment, and considers that other players also view third players' actions as a natural environment, and so on.

Cournot/Nash Equilibrium may provide a theory of cooperation in some special cases that have a unique equilibrium which is Pareto efficient. Otherwise, as Roemer (In Press) argues, attempts to formulate a theory of cooperation based on the Cournot-Nash noncooperative approach are either convoluted, appealing to punishment strategies in indefinitely repeated games, or including, in Roemer's expression, "exotic" factors in addition to the to the  $PMB_i$  in the decision relevant utility function  $U_i$ . We eschew the

---

<sup>4</sup> The consistency of objectives or preferences by itself seems insufficient adequately to characterize rationality: as noted by Amartya Sen (1987, p. 13), "If a person does exactly the opposite of what would help achieving what he or she would want to achieve, and does this with flawless internal consistency (always choosing exactly the opposite of what would enhance the occurrence of things he or she wants and values), the person can scarcely be viewed as rational..."

<sup>5</sup> A parallel formal distinction also appears on the expected utility theory of decision under uncertainty: the von Neumann-Morgenstern-Bernoulli utility function is defined on *outcomes*, whereas the decision-relevant (expected) utility function is defined on *lotteries* or actions.

<sup>6</sup> The literature on *conjectural variations* presents an alternative to the Cournot/Nash equilibrium where a change in an agent's decisions induces changes in the decisions of other agents' decisions (see, e. g., A. L. Bowley (1924), Ragnar Frisch (1951 [1933]), and Timothy Bresnahan (1981).



discussion of indefinitely repeated games, far-removed from our one-shot treatments, and focus on the “exotic” variables approach.

Formally, let there be  $I$  decision makers, indexed  $i = 1, \dots, I$ .  $DM_i$  is a *Cournot/Nash decision maker* if the set of parameters that she faces is the Cartesian product of the set of decisions of the other decision makers:  $P_i = S_1 \times S_2 \dots \times S_{i-1} \times S_{i+1} \dots \times S_I$ . In words, a parameter vector  $p_i$  for her is the vector of (her conjectures on the) decisions of the other decision makers,  $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_I)$ . Her decision-relevant utility function  $U_i$  does not have to coincide with her  $PMB_i$ , but the domain of  $U_i$  is the Cartesian product of the  $S_i$ 's, i. e.,  $U_i: \prod_{i=1}^I S_i \rightarrow \mathfrak{R}$ . Accordingly, given such an  $s_{-i}$ , she chooses  $s_i \in S_i$  in order to maximize  $U_i(s_i; s_{-i})$ .<sup>7</sup>

Explaining the choice of CIRCLE in Table 2 via the Cournot/Nash approach requires rejecting its experimental payoffs as representing the preference relation  $\succsim_i$  on choices. The Cournot/Nash approach could explain the CIRCLE choice there only if the decision-relevant ordering of the decision maker makes CIRCLE better than SQUARE, say because a concern for ethics, joint maximization of payoff gains or fairness, i. e., by appealing to “exotic” arguments in the utility function.

### 3.2.3. Roemer simple Kantian behavior

For the special case where the same set of individual decisions is available to all decision makers, i. e.,  $S_i = S_h \equiv S$ , for all  $i, h = 1, \dots, I$ , a  $DM_i$  as defined in the *Simple Kantian Equilibrium* of Roemer, to be called here a *Kant/Roemer decision maker*, maximizes her personal material benefit  $PMB_i$ , but its domain is effectively restricted to the subset of  $\prod_{i=1}^I S_i$  where all decision makers take the same action, i. e.,

$$U_i: S \rightarrow \mathfrak{R}: U_i(s) = PMB_i(s, \dots, s) .$$

Observe that no parameter appears in the domain of the decision-relevant utility function. We should add that Roemer (In Press) considers an alternative model, called *Conditional Kantian Equilibrium*, where the fraction of other decision makers that choose a cooperative action *is* a parameter in  $U_i$ .

---

<sup>7</sup> It may be the case that the same  $s_i$  maximizes  $U_i$  for all relevant parameter vectors. We call such  $s_i$  a dominant strategy, e. g., SQUARE in the game of Table 2.

We borrow a term from the sciences to denote such an identical action among agents as an *isopraxis*.<sup>8</sup> We can then characterize the behavior of a Kant/Roemer decision maker by two features:

- (1) She chooses among isopraxes only.
- (2) Her decision-relevant utility function  $U_i$  is defined by only her personal material benefit function  $PMB_i$ .

In the example of Table 2, a Kant/Roemer decision maker maximizes her own experimental payoff but considers only the cells in the principal diagonal of the table.

Because of (2), there is no need for the Kant/Roemer decision maker to know the payoff matrix or matrices of the other decision makers: she only has to know that whatever actions are available to her, they are also available to the other decision makers. But the motivation for the Kant/Roemer behavior will suffer when the payoffs are not symmetrical. Section 4.3 below elaborates.

Roemer's Simple Kantian approach does directly justify the CIRCLE choice while keeping the payoffs of Table 2 as representing the relation  $\succsim_i$ . Here the modification affects the "optimization protocol:" instead of parametrically taking the decision of the other player (as in Cournot/Nash optimization), it only looks at the same choices by the two players. The Kant/Roemer decision maker asks herself: do I earn more if we both choose CIRCLE (€10) or if we both choose SQUARE (€ 6)? Because I earn more if we both choose CIRCLE, I choose CIRCLE.

### 3.3. Equilibrium

Both the Cournot/Nash and the Kant/Roemer models are *equilibrium* models: they impose a compatibility condition on the actions of the various decision makers.

Trivially, where the  $U_i$ 's are such that each decision maker has a dominant strategy, Cournot/Nash Equilibrium coincides with Dominant Strategy Equilibrium, and the optimal decisions are not interdependent. The equilibrium condition in the Simple Kantian Equilibrium is one of unanimity: at equilibrium, all Kant/Roemer decision makers choose the same isopraxis. Both in the Dominant Strategy Equilibrium and in the Simple Kantian Equilibrium, the optimal decisions of a decision maker are parameter free, and not interdependent, and equilibrium is reached automatically. A player doesn't have to worry

---

<sup>8</sup> See Paul MacLean (1975).

whether other players play their equilibrium strategies or not, and she doesn't even have to know the payoffs of other players to figure out what they will be playing at equilibrium.

But the optimal decisions of a typical (*i. e.*, excluding the dominant strategy case) Cournot/Nash decision maker are a function of the parametrically taken (conjectured) decisions of other agents, and the equilibrium condition requires such conjectures to be correct. As is frequently the case in equilibrium models, the Cournot/Nash model lacks a convincing formulation of out-of-equilibrium behavior. How is equilibrium reached in this case? If equilibrium is unique, then a common justification is that each player discovers the strategies of other players by solving the game. This of course requires the knowledge of the payoffs of the other players. In addition, if the Cournot-Nash equilibrium is not unique, then a coordination problem appears, which motivates the notion of Correlated Equilibrium (Robert Aumann, 1974).

A conditional Kantian takes as a parameter (her conjecture on) the fraction of other decision makers who behave in the Kantian manner. Accordingly, at a Conditional Kantian Equilibrium such a conjectured fraction must be correct. How is equilibrium reached? If it is reached by player solving the game (unique equilibrium), then she has to know the payoffs of the other players. But it can also be reached by a dynamic adjustment process (unavoidable if equilibrium is not unique), without facing the difficulties of dynamic adjustment process in the Cournot/Nash approach, where the repeated game must be viewed as a larger dynamic game.

#### 4. Roemer Simple Kantian Equilibrium as a theory of cooperation

##### 4.1. Solidarity

As noted above, Roemer's Kant optimization protocol is defined by two properties:

- (1) Restriction to isopraxes.
- (2) Maximization of the  $PMB_i$ .

Features (1) and (2) are independent. For instance,  $DM_i$  may satisfy (1), yet choose among isopraxes according to a criterion other than  $PMB_i$ .

Roemer justifies the Kantian optimization protocol by appealing to *solidarity* (defined as the commonality of interests), and *trust*. In his words (In Press, p. 4):

“Cooperation may be the only means of satisfying one's own self-interested preferences. You and I build a house together so that we may each live in it. We cooperate not because of an interest in the other's welfare, but because cooperative production is the only way of providing any domicile. The same thing is true of the early hunters [...]: without cooperation, neither of us could

capture that deer, which, when caught by our joint effort, will feed both of us. In particular, I cooperate with you because the deer will feed me. It is not necessary that I ascribe any value to the fact it will feed you, too.”

The last sentence evokes the well-known Stag-Hunt game (see, e. g., Brian Skyrms, 2004), as exemplified in Table 4 (CIRCLE corresponds to “STAG” whereas SQUARE corresponds to “HARE” in the common narrative of the game)

|            |        | Column Player |        |
|------------|--------|---------------|--------|
|            |        | CIRCLE        | SQUARE |
| Row Player | CIRCLE | €10           | €1     |
|            | SQUARE | €8            | €6     |

Table 4. Payoff bimatrix of a Stag-Hunt Game

The Simple Kantian equilibrium selects the same pair (CIRCLE, CIRCLE) in both games of Tables 3 and 4. Cournot/Nash, on the contrary, treats the two games quite differently: in Table 3, (SQUARE, SQUARE) is the only equilibrium, actually in the strong form of Dominant Strategy, whereas in Table 4 we have two (pure strategy) Cournot/Nash equilibria: (CIRCLE, CIRCLE) and (SQUARE, SQUARE).

The choice of pair (CIRCLE, CIRCLE) in the Stag Hunt game clearly obeys the principle of solidarity in a stronger sense than in the Prisoner’s Dilemma of Table 3 above. In the Stag-Hunt game of Table 4, (CIRCLE, CIRCLE) Pareto-dominates any other pair of decisions, symmetrical or not, which is not the case in Table 3. We conjecture that if we experimentally tested the Stag-Hunt game (which we did not), we would observe a higher degree of cooperation than what we found (28%) in our Prisoner’s Dilemma experiment.

#### 4.2. Trust

A decision maker playing CIRCLE in the Stag Hunt game of Table 4 is taking a risk, since it would be very bad for her if the other decision maker chose SQUARE. When she chooses CIRCLE, she must have some degree of confidence that the other player is also choosing CIRCLE.

The risk is in a sense higher in the Prisoner’s Dilemma of Table 3, because the other player has so much to gain by playing SQUARE. In fact, Roemer (In Press) does not argue that Kantian behavior is natural in the one-shot Prisoner’s Dilemma:

“Trust, however, must be built up from past experience. I therefore do not claim that it is rational in a truly one-shot game to ask the Kantian question. Nash equilibrium is the rational solution of the truly one-shot game.”

Yet, as noted, 28% of the participants in our one-shot Prisoner Dilemma Treatment did choose the cooperative option CIRCLE, behaving more cooperatively than what Roemer suggests.<sup>9</sup> But the issue of risk was salient in the replies, particularly among the participants who chose SQUARE. In order to address this issue, we performed another treatment, reported in Bosch-Domènech and Silvestre (2017) as the *Riskless and Frameless Prisoner’s Dilemma* Treatment, where the participants were assured that the other player was playing CIRCLE. The fraction of participants who choose the cooperative action CIRCLE then jumped to 49%. This suggests that at least some of the people who would like to cooperate but find it too risky would in fact cooperate if they could trust that the other player was cooperating.

### 4.3 Symmetry

As noted in Section 3.2.3 above, a Kant/Roemer  $DM_i$  only has to know her own  $PMB_i$  and that  $S_i \subset S_h$ , all  $h$ , i. e., the fact that every action that she can take can also be taken by every other decision maker. In fact, Roemer (In Press) defines the Simple Kantian equilibrium for symmetrical games, where payoffs in particular are symmetrical.

Nevertheless, the concept of Simple Kantian equilibrium can well be extended to nonsymmetrical payoffs: equilibrium may not exist there, but it may. Consider for instance the game of Table 5, an asymmetrical variation on the Prisoner’s Dilemma: Kant/Roemer decision makers will still choose the isopraxis (CIRCLE, CIRCLE), making it a simple Kantian equilibrium.

We did not experimentally test the situation of Table 5, but it would be interesting to see whether the fraction of row players who would choose CIRCLE would be higher or lower than what we found for Table 3: an “altruistic” row player may be reinforced to play

---

<sup>9</sup> Of course, it might be argued that our participants in the treatment of Table 3 did not face a “true” one-shot game.

CIRCLE, but now the inequality in the outcomes of (CIRCLE, CIRCLE) may lead a (non Kant/Roemer) row player to prefer (SQUARE, SQUARE) to (CIRCLE, CIRCLE).<sup>10</sup>

|            |        | Column Player |           |
|------------|--------|---------------|-----------|
|            |        | CIRCLE        | SQUARE    |
| Row Player | CIRCLE | €1000<br>€10  | €15<br>€1 |
|            | SQUARE | €1<br>€15     | €6<br>€6  |

Table 5. Payoff bimatrix in asymmetrical variation on Prisoner's Dilemma

#### 4.4. Monotonicity

Why focus on isopraxes? We read in Roemer (In Press, p. 5)

“Assume that it is common knowledge that the payoff matrix is symmetric, and that my opponent and I have equal capacities and reasoning power. Hence, due to the symmetry of the game, whatever strategy I decide upon will also be decided upon by my opponent. It follows that I must only consider strategy profiles  $(x, x)$  as ones that may occur...”

But it seems that something other than symmetry is required for the argument. Consider for instance a game of Chicken, illustrated by Table 6. Depending on the values of  $a$  and  $b$ , it may make sense to focus on cells off the main diagonal, say if  $a = b = 100$ . This could also be interpreted as a situation where efficiency requires the specialization of tasks or the division of labor (a good, balanced meal if either you hunt and I gather, or the other way; but not so good if we both hunt or both gather), so that isopraxes are not efficient.

The simple Kantian equilibrium of the game in Table 6 is, for all values of  $a$  and  $b$ , (CIRCLE, CIRCLE), which for  $a > 10$  and  $b > 10$  is Pareto inferior to either (CIRCLE, SQUARE) and (SQUARE, CIRCLE).<sup>11</sup>

<sup>10</sup> Of course, (SQUARE, SQUARE) is still the only Cournot/Nash equilibrium in Table 5, but it is no longer dominant strategy for the column player.

<sup>11</sup> The Chicken Game is an instance of a symmetrical game where all the pure-strategy Cournot/Nash equilibria are asymmetrical; see Mark Fey (2012) for other examples.

|            |        | Column Player |        |
|------------|--------|---------------|--------|
|            |        | CIRCLE        | SQUARE |
| Row Player | CIRCLE | €10           | € $a$  |
|            | SQUARE | € $b$         | €6     |

Table 6. Payoff bimatrix a Game of Chicken,  $a > 6$ ,  $b > 10$ .

Proposition 2.1 in Roemer (In Press) states that a simple Kantian equilibrium is Pareto efficient if, in addition to symmetry, the game displays monotonicity, which is not the case in the game of Table 6.<sup>12</sup> *Monotonicity* means that

- Either the payoff of each player is increasing in the action(s) of the other player(s) (positive externality)
- Or the payoff of each player is decreasing in the action(s) of the other player(s) (negative externality).

It follows that the Simple Kantian Equilibrium provides a theory of cooperation under externalities, But not if efficiency requires the specialization of tasks.

Of course, Roemer's emphasis on isopraxes is inspired by Kant's Categorical Imperative (Immanuel Kant, 1993 [1785], p. 30)

“Act only according to that maxim whereby you can, at the same time, will that it should become a universal law.”

For instance, consider the following mental experiment:  $DM_i$  lives in a democracy, and a vote is taken on adopting a law that imposes an isopraxis, for instance, mandatory recycling.  $DM_i$  may well vote for making recycling mandatory (which in particular would force  $DM_i$  to recycle), yet in absence of a law she may decide not to recycle, considering that whether she recycles or not makes only an infinitesimal difference. Such a decision maker would not agree with Kant's categorical imperative: if she would like that everybody recycle, then she should recycle whether other people recycle or not.

<sup>12</sup> See also Roemer (In Press, Propositions 3.2 and 3.3), for other results relating efficiency to monotonicity, and Proposition 3.4 therein for a recourse to concavity instead of monotonicity.

But in the statement of the Categorical Imperative the crucial noun is “maxim,” rather than “action.” In the game of Table 6, with  $a = b = 100$ , the Kantian maxim could be “do the opposite of what your opponent does.”

## 5. Conclusion

The experimental comparison between the simple treatments of Tables 1 and 2 yields a powerful conclusion: many decision makers act differently when, *ceteris paribus*, the outcomes of their decisions depend *on an impersonal, natural variable vs. the actions of other human beings*. This validates Roemer’s emphasis on the pervasiveness of cooperation among humans.

We discuss two possible theoretical approaches to model cooperative behavior.

- (1) Cournot/Nash with “exotic” preferences, notably embodying social efficiency or fairness considerations in symmetrical games.
- (2) The Kant/Roemer optimization protocol.

It is difficult experimentally to discriminate between these two approaches. We do not claim to have come close to it in our simple experiments.

Basically, a Cournot/Nash decision maker fails to see other players as capable of joint action or intentionality, and takes the actions of other players as acts of nature. Accordingly, when she makes a cooperative decision she is often unilaterally motivated by considerations of efficiency, fairness or altruism.

In the Kant/Roemer approach, on the contrary, the decision maker explicitly recognizes the social setting within which she decides. She perceives any commonality of interests that may be present, and understands that the other agents do too. The approach develops an elegant unanimity-based equilibrium concept that works well, yielding fairness and Pareto efficiency, in situations with two features:

- (1) Symmetry;
- (2) Monotonicity.

In these situations one can envision the presence of positive or negative externalities. The success of the Kant/Roemer approach under such externalities is substantiated by several appealing applications to general equilibrium models developed in Roemer (In Press), namely:

- Market socialism,
- The production of a public good,



Global carbon emissions,  
Worker-owned firms.

But Roemer's Simple Kantian Equilibrium model does not work as well in the absence of symmetry or of conditions guaranteeing efficiency, such as monotonicity.<sup>13</sup> Assume that the decision sets are identical among decision makers, but that the payoffs are not symmetrical. Then a Simple Kantian Equilibrium may not exist, because Kantian decision makers may choose different isopraxes, violating the unanimity principle.<sup>14</sup> But it may exist in some cases. Yet, if the asymmetries are strong, the Kantian behavior may be harder to justify, casting doubt on the Kant/Roemer restriction to the personal material benefit function ( $PMB_i$ ) as the objective of the decision maker.

The absence of monotonicity, in turn, may lead to situations where the Pareto efficient tuples of decisions are asymmetrical, and therefore neglected by isopraxes-centered Kantian decision makers. A satisfactory theory of cooperation should there focus on these tuples, which may require an explicit appeal to coordination.

Indeed, it can be argued that, in actuality, many instances of cooperation entail coordination via agreements, pacts or norms.<sup>15</sup> But both the Cournot/Nash and the Kant/Roemer theories lack an explicit role for coordination: formally, they provide at best theories of "uncoordinated cooperation." Whereas the lack of coordination may be natural in models of noncooperative interaction, it looks as a limitation in models that attempt to explain cooperation.

---

<sup>13</sup> See footnote 12 above.

<sup>14</sup> This motivates the concepts of multiplicative and the additive Kantian equilibrium, and more generally that of equilibrium relative to a Kantian variation, see Roemer (In Press). These extensions of the simple Kantian equilibrium are in fact the ones utilized in the abovementioned applications to general equilibrium models.

<sup>15</sup> Roemer (In Press) does refer to situations where "trust is warranted either because of past personal experience with potential partners, or because of social convention, of culture."

## APPENDIX: THE TEMPERATURE TREATMENT

We had twenty participants, 16 men and 4 women, all undergraduate students at *Universitat Pompeu Fabra*. We avoided economics and business majors, who may be too familiar with prisoner's dilemma and public goods games, so that most of them majored in law, political science or the humanities.

The experiment took place at the BES Lab of *Universitat Pompeu Fabra*. It was quite simple: participants were individually taken to a separate room, where the instruction sheet, below, was given ("CÍRCULO" is CIRCLE, and "CUADRADO" is SQUARE). After they had made their choice, they were asked to provide an open-format explanation. Once completed, the temperature in Istanbul was checked and the participants paid accordingly.

All twenty participants chose CUADRADO.

### INSTRUCCIONES

Puedes escoger o bien CÍRCULO o bien CUADRADO, pero no ambos. Tus ganancias dependerán de tu decisión y también de la temperatura en Estambul al final del experimento (según CNN Weather). Más exactamente, tus ganancias vienen descritas en el cuadro siguiente.

|                     | Temperatura en Estambul $\leq 16^{\circ}\text{C}$ | Temperatura en Estambul $> 16^{\circ}\text{C}$ |
|---------------------|---|--|
| Tú escoges CÍRCULO  | Tú ganas 10€                                      | Tú ganas 1€                                    |
| Tú escoges CUADRADO | Tú ganas 15€                                      | Tú ganas 6€                                    |

Tus ganancias

**Tú escoges**

**CÍRCULO** \_\_\_\_\_

**CUADRADO** \_\_\_\_\_

(Marca con una cruz lo que hayas escogido)

## REFERENCES

- Aumann, Robert J. (1974). "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics* 1(1), 67-96.
- Bayer, Ralph-C., Elke Renner and Rupert Sausgruber (2013). "Confusion and learning in the voluntary contributions game," *Experimental Economics* 16(4), 478-496.
- Bosch-Domènech, Antoni, and Joaquim Silvestre (2017). "The role of frames, numbers and risk in the frequency of cooperation," *International Review of Economics* 64(3), 245-267.
- Bowley, A. L. (1924). *The Mathematical Groundwork of Economics*. Oxford: Oxford University Press.
- Bowles, Samuel, and Herbert Gintis (2011). *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton: Princeton University Press.
- Bresnahan, Timothy F. (1981). "Duopoly models with consistent conjectures," *American Economic Review* 71, 934-945.
- Burton-Chellew, Maxwell N., Claire El Mouden and Stuart A. West (2016). "Conditional cooperation and confusion in public-goods experiments," *Proceedings of the National Academy of Sciences of the U.S.A.* 113(5), 1291-1296.
- Burton-Chellew, Maxwell N., and Stuart A. West (2013). "Prosocial preferences do not explain human cooperation in public-goods games," *Proceedings of the National Academy of Sciences of the U.S.A.* 110(1), 216-221.
- Clark, Kenneth, and Martin Sefton (2001). "The sequential Prisoner's Dilemma: Evidence on reciprocation," *Economic Journal* 111(468), 51-68.
- Cooper, David J., and John H. Kagel (2015). "Other-regarding preferences: A selective survey of experimental results," in John H. Kagel and Alvin E. Roth, Editors, *Handbook of Experimental Economics, Vol. 2*. Princeton, NJ: Princeton University Press.
- Cournot, Antoine Augustin (1838). *Recherches sur les principes mathématiques de la théorie des richesses*. Paris: Hachette.
- Fey, Mark (2012). "Symmetric games with only asymmetric equilibria," *Games and Economic Behavior* 75(1), 424-427.
- Frisch, Ragnar (1951 [1933]). "Monopoly – Polypoly – The concept of force in the economy," *International Economic Papers* 1, 23-36.

- Houser, Daniel, and Robert Kurzban (2002). "Revisiting kindness and confusion in public goods experiments," *American Economic Review* 92(4), 1062–1069.
- Kant, Immanuel (1993) [1785]. *Grounding for the Metaphysics of Morals*. Translated by Ellington, James W. (3<sup>rd</sup> ed.). Hackett.
- Kümmerli, Rolf, Maxwell N. Burton-Chellew, Adin Ross-Gillespie and Stuart A. West (2010). "Resistance to extreme strategies, rather than prosocial preferences, can explain human cooperation in public goods games," *Proceedings of the National Academy of Sciences of the U.S.A.* 107(22), 10125–10130.
- MacLean, Paul D. (1975). "The imitative-creative interplay of our three mentalities," in Harold Harris, Editor, *Astride the Two Cultures. Arthur Koestler at 70*. London: Hutchinson.
- Mas-Colell, Andreu, Michael D. Whinston and Jerry R. Green (1995). *Microeconomic Theory*. Oxford: Oxford University Press.
- Nash, John F. (1950). "Equilibrium points in N-person games," *Proceedings of the National Academy of Sciences of the U. S. A.* 36 (1), 48–49.
- Nash, John F. (1951). "Non-cooperative games," *Annals of Mathematics* 54 (2), 286–95.
- Roemer, John E. (2010). "Kantian equilibrium," *Scandinavian Journal of Economics* 112, 1-24.
- Roemer, John E. (2015). "Kantian optimization: A microfoundation for cooperation," *Journal of Public Economics* 127, 45-57.
- Roemer, John E. (2017a) "A theory of cooperation with application to market socialism," unpublished.
- Sen, Amartya (1987). *On Ethics & Economics*. Oxford: Blackwell.
- Skyrms, Brian (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.